

# 1 Storingsgevoeligheid van netwerken

Floske Spieksma, Universiteit Leiden

## Inleiding

Hoe gevoelig is een trein- of metronetwerk voor grote oponthouden als er op een traject een incident gebeurt? Zijn er voldoende omleidingsmogelijkheden, of moeten reizigers gestrand op een station kamperen...

De storingsgevoeligheid van twee netwerken kan alleen zinvol vergeleken worden, als er een maat is die ‘storingsgevoeligheid’ kwantificeert. Een metronetwerk kan gemodelleerd worden als een collectie stations (‘knooppunten’) en trajecten tussen stations (‘takken tussen knooppunten’) oftewel een ‘graaf’. Een lage storingsgevoeligheid betekent een hoge ‘robuustheid’.

We zullen een aantal eigenschappen bespreken waaraan zinvolle maten moeten voldoen. Daarnaast laten we een aantal robuustheidsmaten de revue passeren, met name de maat die ‘effectieve weerstand’ heet.

Deze maat is gebaseerd op de weerstand in elektrische netwerken, en heeft de interessante eigenschap dat zij niet alleen via de wetten van Kirchhoff en Ohm, maar ook op minstens zes andere manieren berekend kan worden: onder andere aan de hand van opspannende bomen van een graaf, of via de trajectduur tussen twee willekeurige steden van een random rijdende trein.

Interpreteer nu steden als mensen en een random rijdende trein als een verstuurd wordende kettingbrief. Dan is de trajectduur van een random rijdende trein vanaf een willekeurige stad naar Amsterdam niet anders dan de duur die een kettingbrief erover doet om de Paus te bereiken, wanneer de kettingbrief steeds naar een willekeurige bekende wordt doorgestuurd.

Dat laatste doet denken aan de bekende ‘six degrees of separation’ uitspraak: de gemiddelde afstand tussen U en de Paus is hooguit zes. Wat betekent dat? De kettingbrief is gebruikt in een experiment in de 60’er jaren (als eerste van vele) om de ‘six degrees of separation’ hypothese te testen.. Klopt dit eigenlijk wel, hoe kunnen we dit inzien en wat is de relatie met ‘robuustheid van een netwerk’?

# 1.1 Gemiddelde afstand en kortste paden

Grafen zijn een standaard representatie van netwerken als een wegnetwerk (zie Fig. 1.1). Ook sociale netwerken, en het collectivum van onderlinge links tussen webpagina's kan men hiermee modelleren.

Formeel bestaat een graaf  $G = (V, E)$  uit een eindige niet-lege verzameling  $V$ , de *knooppunten*, en een niet-geordende verzameling  $E$  van tweetallen uit  $V$ , de *takken*. Het aantal knooppunten, respectievelijk het aantal takken, geven we aan met  $n$ , respectievelijk  $m$ .

Als  $e = (i, j) \in E$ , dan zeggen we dat  $e$  *incident* is met  $i$  en met  $j$ . De graad  $d(i)$  van knooppunt  $i \in V$  is het aantal takken dat incident is met  $i$ , en  $N(i) \subset V$  is de collectie knooppunten die via een tak met knooppunt  $i$  verbonden is.

Een *takkenreeks* is een rij takken, waarbij elk tweetal opeenvolgende takken een knooppunt gemeen heeft. De takkenreeks heet een *keten* als elke tak hooguit één keer in de takkenreeks voorkomt.

Een belangrijk concept is *samenhang*: een graaf  $G$  heet *samenhangend* als er tussen elk tweetal knooppunten een keten in  $G$  is. Een samenhangende graaf heet een *boom* als er tussen elk knooppunten precies één keten is, oftewel,  $m = n - 1$ . Omdat robuustheid een maat voor de onderlinge verbondenheid van het netwerk, of de graaf, is, zullen we slechts samenhangende grafen bestuderen.

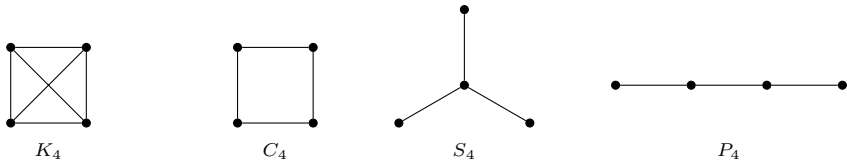
Een goede robuustheidsmaat moet voldoen aan de volgende eigenschappen:

- de maat moet gebaseerd zijn op eigenschappen die gerelateerd zijn aan onze intuïtie met betrekking tot robuustheid;
- de maat moet gemakkelijk te berekenen zijn;
- bij een toenemend aantal takken mag de robuustheid niet afnemen.

Laten we de volgende grafen eens bekijken:



Figuur 1.1: Metronet Amsterdam



Van links naar rechts worden de grafen respectievelijk de volledige graaf  $K_4$ , de cykelgraaf  $C_4$ , de stergraaf  $S_4$  en de padgraaf  $P_4$  genoemd. ‘Volledig’ betekent dat er door toevoeging van takken parallelle takken ontstaan of lussen. De index 4 slaat op het aantal knooppunten.

Achtereenvolgens worden deze grafen gevoeliger voor uitval van takken, ze zijn ‘minder verbonden’ en derhalve moet een goede robuustheidsmaat de grafen in die volgorde lager waarderen.

Verbondenheid kan gekwantificeerd worden via de afstand tussen een willekeurige paar knooppunten. Hoe korter de afstand tussen twee knooppunten des te ongevoeliger de verbinding is voor storingen.

Zij  $G = (V, E)$  een samenhangende graaf. De *afstand*  $d_{ij}$  tussen knooppunten  $i$  en  $j$  is het aantal takken in de kortste keten tussen  $i$  en  $j$ . De *gemiddelde afstand*  $\bar{d}$  is dan

$$\bar{d} = \frac{2 \sum_{i=1}^n \sum_{j=i+1}^n d_{ij}}{n(n-1)},$$

met andere woorden, de som van alle afstanden gedeeld door het maximale aantal takken in een volledige graaf.

De maat  $\bar{d}$  is in het algemeen niets anders dan de ‘degree of separation’ in een sociale netwerk graaf. Hierin worden de mensen uit het sociale netwerk gepresenteerd als knooppunten. De aanwezigheid van een tak representeert dat ‘incidente’ mensen elkaar kennen, een soort ‘kennis’-graaf dus! We schrijven ‘in het algemeen’ omdat er geen eenduidige definitie van het aantal ‘degrees of separation’ lijkt te zijn.



Soms wordt er met ‘degrees of separation’ bedoeld: ‘de maximale afstand die zich voordoet in het genoemde sociale netwerk’. Dat deze hooguit zes zou zijn, is overigens natuurlijk onzin, wanneer we alle mensen op Aarde beschouwen! Wij houden hier dus op  $\bar{d}$ !

Een tweede maat is gekoppeld aan de takken: de *takspreiding* (edge betweenness). Dit is een gewogen som van het aantal kortste ketens tussen elk knooppuntenpaar dat door tak  $e$  gaat. Zij  $n_{ij}(e)$  het aantal kortste ketens tussen knooppunten  $i$  en  $j$  dat door tak  $e$  gaat, en  $n_{ij}$  het aantal kortste ketens tussen  $i$  en  $j$ . Dan is  $n_{ij}(e)/n_{ij}$  de fractie van kortste ketens tussen  $i$  en  $j$  die door tak  $e$  gaat. De robuustheid is slechter naarmate deze fractie groter is. Immers, het aantal alternatieven voor routes door  $e$  is dan kleiner.

Als een graaf een verkeersnetwerk modelleert, waarbij de hoeveelheid verkeer over elke tak gelijk is, en verkeer altijd de kortste weg (keten) neemt, dan kunnen we  $\sum_{j>i} n_{ij}(e)/n_{ij}$  als de belasting van tak  $e$  interpreteren. Hiermee kunnen file-gevoelige wegdelen worden geïdentificeerd.

We definiëren de *takspreiding*  $\bar{n}(e)$  van tak  $e$  als

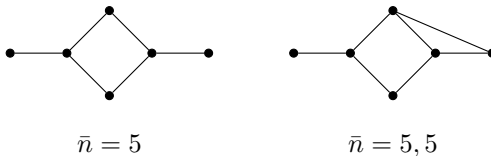
$$\bar{n}(e) = \sum_{j>i}^n \frac{n_{ij}(e)}{n_{ij}}.$$

Dan is  $\bar{n} = \max_e \bar{n}(e)$  de takspreiding van de graaf. Girvan en Newman [2] gebruikten de takverbondenheid om sociale subgemeenschappen te identificeren. Het idee is dat takken met een hoge waarde  $\bar{n}(e)$  corresponderen met de *relaties tussen* de verschillende gemeenschappen.

Voor de 4 grafen zijn de gemiddelde afstand en takspreiding als volgt:

	$K_4$	$C_4$	$S_4$	$P_4$
$\bar{d}$	1	4/3	3/2	5/3
$\bar{n}$	1	2	3	4

De uitkomsten van deze twee maten komen overeen met onze intuïtie over de netwerken. De gemiddelde afstand neemt af door toevoeging van takken. Het volgende voorbeeld laat zien dat de takspreiding niet noodzakelijk afneemt.



**Opgave 1.1.1.** Ga na dat de tweede graaf inderdaad een grotere takspreiding heeft.

De maat ‘takspreiding’ voldoet dus niet aan de eisen voor een robuustheidsmaat. De ‘gemiddelde afstand’ heeft dit nadeel niet.

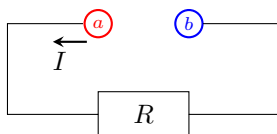
Hij heeft wel het nadeel dat alleen kortste ketens worden meegewogen, en niet het aantal alternatieve routes. Verder is de gemiddelde afstand niet gemakkelijk te berekenen bij grote grafen. We zullen daarom een alternatieve maat beschouwen.

Alvorens hiertoe over te gaan, zult U zich afvragen of we niet ook de daadwerkelijke weglengtes zouden moeten meemodelleren. Dat kan inderdaad, zie onderstaande opgave.

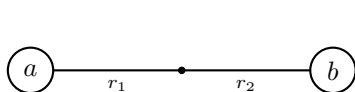
**Opgave 1.1.2.** *Stel dat de lengte van tak  $(i, j)$  niet gelijk is aan 1, maar lengte  $l(i, j)$  heeft. Hoe moet de formule voor de gemiddelde afstand dan worden aangepast? Stel vervolgens dat we de ‘verkeersbelasting’ van wegen in een wegennetwerk willen meemodelleren. Uitval van een wegdeel (tussen twee knooppunten) met een hoge verkeersbelasting wordt als schadelijker ervaren dan uitval van een wegdeel met een lage verkeersbelasting. Pas de formule voor de gemiddelde afstand hierop aan.*

## 1.2 Electriche netwerken

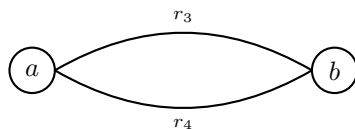
We zullen nu een robuustheidsmaat introduceren die gebaseerd is op elektrische netwerken. Het idee is dat bij een vast spanningsverschil een toenemende weerstand voor een lagere stroom zorgt. Een lagere stroom kunnen we opvatten als een slechtere verbondenheid.



Voor het berekenen van de weerstand tussen twee knooppunten van een elektrisch netwerk, kunnen we de bekende regels voor serie- en parallelschakeling toepassen. De waarden van de vervangingsweerstand tussen knooppunten  $a$  en  $b$ , oftewel de *effectieve weerstand*  $R_{ab}$ , staan in de onderstaande figuren.



$$R_{ab} = r_1 + r_2$$



$$R_{ab} = \frac{1}{\frac{1}{r_3} + \frac{1}{r_4}}$$

**Opgave 1.2.1.** *Beschouw het deel van Amsterdamse metronetwerk tussen Centraal Station via Spaklerweg naar Zuid. Dit bevat delen van de groene, gele, rode en oranje lijnen, maar niet van de blauwe. Modelleer dit als een elektrisch netwerk met een spanningsbron tussen Zuid en Centraal, waarbij aan elke tak van het metronetwerk een weerstand van 1 Ohm is toegekend. Bereken de effectieve weerstand  $R_{\text{Centraal, Zuid}}$ .*

U ziet dat dit niet zonder meer te generaliseren is naar een graaf. Hoe zouden we überhaupt de weerstanden in de  $K_4$  en zelfs de  $K_5$  moeten berekenen?

We kunnen dit oplossen met behulp van de stroomwet van Kirchhoff (stroom in = stroom uit, in elk knooppunt ongelijk aan de spanningsbron) en de wet van Ohm ( $V = IR$ ).

Beschouw een elektrisch netwerk met spanningsbron tussen  $a$  en  $b$ , gepresenteerd als een graaf. Er loopt een stroom ter grootte van 1 Ampère van  $a$  naar  $b$ . De weerstand tussen twee aangrenzende knooppunten  $i$  en  $j$  is  $r_{ij}$ , en de stroom door de tak van  $i$  naar  $j$  is  $y_{ij}$  (N.B.  $y_{ij} = -y_{ji}$ !). Van de stroomgroottes weten we op dit moment slechts dat  $\sum_{j \in N(a)} y_{aj} = 1$  en  $\sum_{j \in N(b)} y_{bj} = -1$ .

Volgens de wet van Ohm is het spanningsverschil tussen  $i$  en  $j$  gegeven door:  $v_i - v_j := r_{ij}y_{ij}$ . Volgens de stroomwet van Kirchhoff geldt dat

$$\sum_{j \in N(i)} \frac{v_i - v_j}{r_{ij}} = \sum_{j \in N(i)} y_{ij} = \begin{cases} 1, & i = a \\ 0, & i \neq a, b \\ -1, & i = b. \end{cases} \quad (1.1)$$

Oplossen van dit stelsel geeft de spanningen  $\mathbf{v} = (v_i)_{i \in V}$ . Volgens de wet van Ohm, geldt  $v_a - v_b = R_{ab} \cdot 1 = R_{ab}$ . Kortgezegd, vervanging van het hele netwerk door één tak tussen  $a$  en  $b$  met vervangingsweerstand  $R_{ab}$ , geeft het gevraagde spanningsverschil.

Bovenstaand lineair stelsel vergelijkingen kunnen we nu gebruiken om de effectieve weerstand te definiëren voor een samenhangende graaf  $G = (V, E)$ . Neem aan, dat er geen parallelle takken en lussen zijn. Geef elke tak een weerstand 1 (Ohm):  $r_{ij} = 1, (i, j) \in E$ . Definieer de  $n \times n$  Laplaciaan matrix  $L_G$  als volgt:

$$L_G(i, j) = \begin{cases} d(i), & i = j \\ -1, & j \neq i \\ 0, & \text{anders.} \end{cases}$$

Voor de  $S_4$  krijgen we dan (met knooppunt 1 als het middelste knooppunt)

$$\mathbf{L}_{S_4} = \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 1 & & \\ -1 & & 1 & \\ -1 & & & 1 \end{pmatrix},$$

waarbij de 0-en zijn weggelaten. Met  $\mathbf{v} = (v_i)_{i \in V}$  reduceert stelsel (1.1) tot

$$\mathbf{L}_G \mathbf{v} = \mathbf{e}_a - \mathbf{e}_b, \quad (1.2)$$

met  $\mathbf{e}_i$  de  $n$ -dimensionale (kolom)vector met  $i$ -de component gelijk aan 1 en de rest 0.

**Opgave 1.2.2.** *Verifieer dit.*

Schrijf  $\mathbf{1}_n$  voor de  $n$ -dimensionale (kolom)vector bestaande uit alleen 1-en, en  $\mathbf{0}_n$  voor de  $n$ -dimensionale (kolom)vector bestaande uit alleen 0-en. Dan geldt

$$\mathbf{L}_G \mathbf{1}_n = \mathbf{0}_n,$$

zodat  $\mathbf{1}_n$  een (rechter-)eigenvector bij eigenwaarde 0 is. De matrix  $\mathbf{L}_G$  is dientengevolge singulier, en vergelijking (1.2) niet uniek oplosbaar is, mits er überhaupt een oplossing is.

Het kan bewezen worden dat in een samenhangende graaf de dimensie van de nulruimte van  $\mathbf{L}_G$  gelijk is 1. In het onderhavige geval is de nulruimte  $\{\alpha \mathbf{1} \mid \alpha \in \mathbf{R}\}$ .

Aangezien  $\mathbf{L}_G$  een symmetrische matrix is, heeft  $\mathbf{L}_G$  dientengevolge een orthogonale basis van eigenvectoren. Omdat  $\mathbf{e}_a - \mathbf{e}_b \perp \mathbf{1}$ , is  $\mathbf{e}_a - \mathbf{e}_b$  een lineaire combinatie van de eigenvectoren ongelijk  $\mathbf{1}_n$ . Dit impliceert dat vergelijking (1.2) ook minstens één oplossing heeft.

Uit bovenstaande volgt onmiddellijk dat het verschil  $v_a - v_b$  onafhankelijk is van de gekozen oplossing  $\mathbf{v}$  van (1.2). Elke willekeurige oplossing  $\mathbf{v}$  van vergelijking (1.2) voldoet.

**Definitie 1.2.1.** *Laat  $G = (V, E)$  een samenhangende graaf zijn, zonder parallele takken en lussen en laat  $\mathbf{v}$  een oplossing van vergelijking (1.2) zijn.*

- $R_{ab} = v_a - v_b$  is effectieve weerstand tussen twee knopen  $a, b$ .
- $R^{tot} = \sum_{j>i} R_{ij}$  is de totale effectieve weerstand.

De totale effectieve weerstand zullen we als maat gebruiken voor de robuustheid van een netwerk: een hogere weerstand staat voor een lagere robuustheid.

Daartoe moeten we nog wel nagaan of deze voldoet aan de eigenschappen die we hadden gepostuleerd. Op het eerste gezicht lijkt verder de totale effectieve weerstand niet gemakkelijk te berekenen: per paar  $(a, b)$  lijken we een singulier stelsel vergelijkingen te moeten oplossen. We zullen zien dat dat meevalt.

### 1.3 Berekening $R_{ab}$ en $R^{tot}$ via vegen

De centrale uitspraak waarmee de effectieve weerstanden berekend kunnen worden is de volgende. Eerst weer enige notatie:  $\mathbf{L}_{G,[i]}$  is de matrix waarbij de  $i$ -de rij en kolom van  $\mathbf{L}_G$  zijn weggelaten. De matrix  $\mathbf{L}_{G,[i,j]}$  is de matrix waarbij  $i$ -de en  $j$ -de rij en kolom zijn weggelaten. De getransponeerde van een (kolom)vector  $\mathbf{v}$  is de rijvector  $\mathbf{v}'$ .

**Stelling 1.3.1.** *Kies  $b \in V$  willekeurig. Dan is  $\mathbf{L}_{G,[b]}$  inverteerbaar met inverse  $\mathbf{L}_{G,[b]}^{-1}$ , en er geldt:*

$$R_{ij} = \begin{cases} \mathbf{L}_{G,[b]}^{-1}(i, i), & j = b, i \neq b \\ \mathbf{L}_{G,[b]}^{-1}(j, j), & j \neq b, i = b \\ \mathbf{L}_{G,[b]}^{-1}(i, i) + \mathbf{L}_{G,[b]}^{-1}(j, j) - \mathbf{L}_{G,[b]}^{-1}(i, j) - \mathbf{L}_{G,[b]}^{-1}(j, i), & i, j \neq b \end{cases}$$

en

$$R^{tot} = n \sum_i \mathbf{L}_{G,[b]}^{-1}(i, i) - \mathbf{1}'_{n-1} \mathbf{L}_{G,[b]}^{-1} \mathbf{1}_{n-1}.$$

Merk op dat de effectieve weerstanden symmetrisch zijn:  $R_{ij} = R_{ji}$ !

Laten we de effectieve weerstanden berekenen voor de stergraaf  $S_3$ . Kies  $b = 1$ , met knooppunt 1 het centrale knooppunt. Dan volgt

$$\mathbf{L}_{S_4,[1]} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix},$$

en dus geldt  $\mathbf{L}_{S_4,[1]}^{-1} = \mathbf{L}_{S_4,[1]}$ , zodat  $R^{tot} = 4 \cdot 3 - 3 = 9$ .

Het lijkt vreemd dat dit onafhankelijk is van het gekozen uitgezonderde knooppunt. Voor wie het niet gelooft, (en om U daarmee het vegen weer in herinnering te brengen), zullen we dit nogmaals doen, maar dan met



$b = 4$ :

$$L_{S_4, [4]} = \begin{pmatrix} 3 & -1 & -1 \\ -1 & 1 & \\ -1 & & 1 \end{pmatrix},$$

Vegen houdt in dat door rijoperaties de matrices  $L_{S_4, [4]}$  en de identiteit  $I$  simultaan naar  $I$  en de inverse  $L_{S_4, [4]}^{-1}$  worden getransformeerd:

$$\left( \begin{array}{ccc|cc} 3 & -1 & -1 & 1 & \\ -1 & 1 & & & 1 \\ -1 & & 1 & & 1 \end{array} \right)$$

Tel  $1/3$  keer de eerste rij bij de twee volgende rijen op:

$$\left( \begin{array}{ccc|cc} 3 & -1 & -1 & 1 & \\ & 2/3 & -1/3 & 1/3 & 1 \\ & -1/3 & 2/3 & 1/3 & 1 \end{array} \right)$$

Tel nu  $1/2$  keer de tweede rij bij de derde op:

$$\left( \begin{array}{ccc|ccc} 3 & -1 & -1 & 1 & & \\ & 2/3 & -1/3 & 1/3 & 1 & \\ & & 1/2 & 1/2 & 1/2 & 1 \end{array} \right).$$

We maken nullen in de derde kolom door twee keer de derde rij bij de eerste op te tellen en  $2/3$ -de keer bij de tweede:

$$\left( \begin{array}{ccc|ccc} 3 & -1 & & 2 & 1 & 2 \\ & 2/3 & & 2/3 & 4/3 & 2/3 \\ & & 1/2 & 1/2 & 1/2 & 1 \end{array} \right)$$

Tel nu  $3/2$ -de keer de tweede rij bij de eerste op:

$$\left( \begin{array}{ccc|ccc} 3 & & & 3 & 3 & 3 \\ & 2/3 & & 2/3 & 4/3 & 2/3 \\ & & 1/2 & 1/2 & 1/2 & 1 \end{array} \right)$$

Nu nog de eerste rij met  $1/3$  vermenigvuldigen, de tweede met  $3/2$  en de derde met  $2$ :

$$\left( \begin{array}{ccc|ccc} 1 & & & 1 & 1 & 1 \\ & 1 & & 1 & 2 & 1 \\ & & 1 & 1 & 1 & 2 \end{array} \right),$$

zodat

$$L_{S_4, [4]}^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

Inderdaad geldt weer:  $R^{\text{tot}} = 9!$

**Bewijs van Stelling 1.3.1** Zonder verlies van algemeenheid kunnen we knooppunt  $b$  als het hoogste genummerde knooppunt  $n$  kiezen. Pas een veegprocedure op  $(L_G | I)$  toe, waar bij als eerste stap de eerste  $n - 1$  rijen bij de laatste worden opgeteld. Om de rijsummen van  $L_G$  de nul-vector opleveren, bestaat de laatste rij uit

$$(\mathbf{0}_n | \mathbf{1}_n),$$

en daar gebeurt verder niets meer mee. Doorvegen op de eerste  $n - 1$  rijen levert uiteindelijk  $(T|S)$  op, met

$$T = \begin{pmatrix} 1 & & & -1 \\ & 1 & & -1 \\ & & \ddots & -1 \\ & & & 1 & -1 \\ & & & & 1 - 1 \end{pmatrix} \quad \text{en} \quad S = \begin{pmatrix} L_{G,[n]} & \mathbf{0}_{n-1} \\ \mathbf{1}_{n-1} & 1 \end{pmatrix}.$$

De laatste kolom van  $T$  kan worden verklaard uit het feit de  $L_G$  kolomsommen 0 heeft, en die eigenschap verandert niet door vege. De rij-operaties die  $T$  uit  $L_G$  genereren, maken de identiteit  $I$  tot  $S$ , hetgeen impliceert dat  $SL_G = T$ .

Bekijk de vergelijking  $L_G x = y$ . Er geldt:  $SL_G x = Tx$ , oftewel

$$Sy = Tx = x - x_n \mathbf{1}_n.$$

Laat  $a, b \in V$ , en vul  $y = \mathbf{e}_a - \mathbf{e}_b$  in. Dan geldt

$$x_a - x_b = S(\mathbf{e}_a - \mathbf{e}_b)_a - S(\mathbf{e}_a - \mathbf{e}_b)_b.$$

De vorm van  $S$  in aanmerking nemende, volgt de uitspraak van de stelling direct.  $\square$

**Opgave 1.3.1.** 1. Bereken de effectieve weerstanden en de totale effectieve weerstand voor de padgraaf  $P_4$ .

2. Vergelijk  $R_{a,b}$  met de afstand  $d_{a,b}$ , en  $R^{\text{tot}}$  met  $\bar{d}$ .

3. Doe dezelfde vergelijking maar dan voor de stergraaf  $S_4$ .

4. Kunt  $U$  een algemene uitspraak hierover formuleren?

Zoals eerder opgemerkt, kan de modelleerder het belangrijk vinden om weglengtes in wegnennetwerk mee te modelleren, of de verkeersbelasting. Ook zal  $U$  vast niet ontgaan zijn, dat aan de eis dat er geen parallele takken zijn door het Amsterdamse metronetwerk niet wordt voldaan.

**Opgave 1.3.2.** • *Opgave 1.1.2 met ‘gemiddelde afstand’ vervangen door ‘effectieve weerstand’. Dit komt neer op het definiëren van geschikte weerstanden op de takken en het bouwen van een equivalent van de Laplaciaan matrix (de zogenaamde gewogen Laplaciaan matrix). Beschouw voor Uw ideevorming eerst het stuk van het Amsterdamse metronetwerk tussen Centraal en Zuid, maar nu inclusief de blauwe lijn.*

• *Geldt de uitspraak van Stelling 1.3.1 met een gewijzigde Laplaciaan?*

Ik hoop dat ik U ervan heb kunnen overtuigen dat de effectieve weerstand gemakkelijk te berekenen is... Ook numeriek is deze matrix goed te inverteren.

We zijn nog niet ingegaan op de vraag, of de effectieve weerstand wel aan de derde eis voldoet, namelijk dat deze niet toeneemt na toevoeging van takken.

**Lemma 1.3.2.** *Bij toevoeging van een tak tussen knooppunt  $a$  en  $b$ , neemt de effectieve weerstand tussen elk paar knooppunten niet toe.*

*Deelbewijs Lemma 1.3.2.* Beschouw de matrix  $L_{G,[b]}$ . Door toevoeging van de tak  $(a, b)$  verandert slechts het element  $L_{G,[b]}(a, a)$ . Via vegen kun je aantonen dat  $R_{ij}$  niet toeneemt,  $j \neq i$ .

**Opgave 1.3.3.** *Bewijs Lemma 1.3.2.*

In de volgende twee paragrafen gaan we in op alternatieve interpretaties van de effectieve weerstand.

## 1.4 Gemiddelde afstand via een random surfer

Gegeven een samenhangende graaf. Een random surfer surft van knooppunt naar knooppunt: op het moment dat hij in knooppunt  $i$  is, kiest hij een volgend knooppunt  $j \in N(i)$  met kans  $1/d(i)$ . Die kansen zijn dus afhankelijk van het knooppunt waarin de surfer zich bevindt, en die kunnen we daarom het handigst in een matrix representeren, die we  $P_G$  noemen:

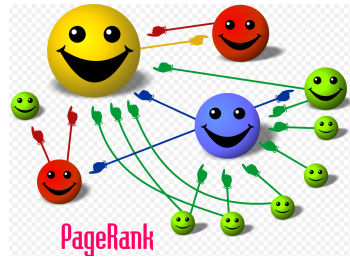
$$P_G(i, j) = \frac{1}{d(i)}, \quad j \in N(i), j \neq i, \quad i \in V.$$

Voor de stergraaf  $S_4$  is dit

$$P_{S_4} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Degenen onder U die zich met Markovketens hebben beziggehouden herkennen hierin de transitiematrix van een Markovketen.

Hoe kan surfen leiden tot een formule voor de effectieve weerstand? Analoog aan hoe surfen leidt tot een waarde voor Google PageRank. De webgraaf is de graaf met als knooppunten *alle* webpagina's (bijna 4,5 miljard.), waar bij er een gerichte tak  $(i, j)$  als er een link op pagina  $i$  naar pagina  $j$  is.



De Google PageRank van pagina  $i$  is in feite de fractie van het aantal bezoeken van een random surfer aan pagina  $i$  (modulo nog enkel details en onbekende trucs die Google langzamerhand heeft ingebouwd).

Laten we de Laplaciaan eens vergelijken met  $P_G$ . De volgende relatie geldt:

$$D(\mathbf{I} - P_G) = L_G,$$

waarbij  $D$  een diagonaal matrix is met  $D(i, i) = d(i)$ . Evenzo geldt een analoge formule voor  $L_{G,[b]}$ :

$$D_{[b]}(\mathbf{I} - P_{G,[b]}) = L_{G,[b]}.$$

Voor  $S_4$  geldt dat  $P_{S_4,[1]}$  de nulmatrix is. In het algemeen zijn de rijssommen van  $P_{G,[b]}$  niet alle meer gelijk aan 1, en wegens de samenhang van de graaf impliceert dit dat de inverse matrix  $(\mathbf{I} - P_{G,[b]})$  bestaat en gelijk is aan

$$(\mathbf{I} - P_{G,[b]})^{-1} = \sum_{k=0}^{\infty} P_{G,[b]}^k,$$

waarbij  $P_{G,[b]}^k$  de  $k$ -de macht van de matrix is. N.B. dit is geheel analoog aan de gelijkheid  $\sum_{k=0}^{\infty} x^k = (1 - x)^{-1}$  voor  $|x| < 1$ . Dus geldt

$$L_{G,[b]}^{-1} = \sum_{k=0}^{\infty} P_{G,[b]}^k D_{[b]}^{-1}.$$

We zullen nu een interpretatie aan de som in het rechterlid geven.

Door weglaten van de rij en kolom van  $b$ , nemen we de surfer slechts waar zolang hij nog niet in  $b$  geweest. Dit komt erop naar dat de surfer doorsurft tot hij  $b$  bereikt heeft. De interpretatie van het element  $P_G^k(i, j)$

is de kans dat de ophoudende random surfer na  $k$  keer surfen in knooppunt  $j$  is bij start in  $i$ . De oneindige som geeft dan precies het verwachte aantal keren dat de surfer in knooppunt  $j$  is geweest, bij start in  $i$ , alvorens in  $b$  aan te landen. Laten we dit noteren met  $\tau_{[b]}(i, j)$ . Noteer met  $\tau(i, b)$  het verwachte aantal keren dat de surfer moet surfen om  $b$  te bereiken.

Er geldt dat

$$\tau(i, b) = \sum_{j \neq b} \tau_{[b]}(i, j). \quad (1.3)$$

Dit mag U zelf eens proberen te beargumenteren! Hieruit volgt dat

$$\mathbb{L}_{G, [b]}^{-1}(i, j) = \tau_{[b]}(i, j) \frac{1}{d(j)}, \quad i, j \neq b. \quad (1.4)$$

De volgende verrassende stelling geldt nu.

**Stelling 1.4.1.**  $R_{i,j} = \frac{\tau(i, j) + \tau(j, i)}{\sum_{l=1}^n d(l)}$ .

Een hoge weerstand betekent dus dat de surfer er lang over doet om van  $i$  naar  $j$  te komen en/of vice versa.

Alvorens deze uitspraak te bewijzen, kunnen we met behulp van vergelijkingen (1.3) en (1.4) nog iets zeggen over het effect van het toevoegen van een tak  $(a, b)$  op de effectieve weerstand met een random surfer argument.

*Vervolg deelbewijs Lemma 1.3.2* Na toevoeging van de tak  $(a, b)$ , springt de surfer met grotere kans onmiddellijk van  $a$  naar  $b$  (en omgekeerd), en met kleinere kans naar een ander aangrenzend knooppunt. U zult direct geloven dat  $\tau(i, b)$  hierdoor niet toeneemt,  $i \neq a, b$ . Immers, elke keer dat de surfer in  $a$  komt, springt hij met grotere kans bij de volgende surf naar  $b$ , en dan houdt hij op met surfen.

**Opgave 1.4.1.** 1. Geef een formule voor  $\tau(i, b)$  in termen van de inverse  $\mathbb{L}_{G, [b]}^{-1}$ .

2. Gebruik de symmetrie van  $\mathbb{L}_{G, [b]}^{-1}$  om een relatie af te leiden tussen  $\tau_{[b]}(i, j)$  en  $\tau_{[b]}(j, i)$ .

3. Toon aan dat  $R_{i,b}$  niet toeneemt door toevoeging van de tak  $(a, b)$ , voor  $i \neq a, b$ , met een random surfer argument.

**Bewijs Stelling 1.4.1** Laat  $i, j \in V$ ,  $i \neq j$ . Dan geldt

$$\begin{aligned} \tau(i, j) &= \sum_k \tau_{[j]}(i, k) \\ &= \sum_k \mathbb{L}_{G, [j]}^{-1}(i, k) d(k). \end{aligned} \quad (1.5)$$

Nu geldt voor  $k \neq i$ , wegens de symmetrie van de inverse,

$$R_{ik} = \mathbb{L}_{G,[j]}^{-1}(i, i) + \mathbb{L}_{G,[j]}^{-1}(k, k) - 2\mathbb{L}_{G,[j]}^{-1}(i, k),$$

zodat

$$\mathbb{L}_{G,[j]}^{-1}(i, k) = \frac{1}{2}(\mathbb{L}_{G,[j]}^{-1}(i, i) + \mathbb{L}_{G,[j]}^{-1}(k, k) - R_{ik}) = \frac{1}{2}(R_{ij} + R_{kj} - R_{ik}).$$

We vullen dit in vergelijking (1.5) in:

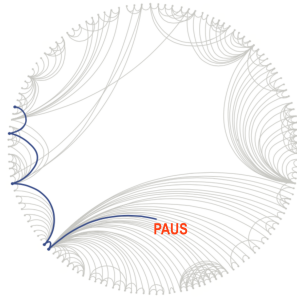
$$\tau(i, j) = d(i)R_{ij} + \frac{1}{2} \sum_{k \neq i, j} d(k)(R_{ij} + R_{kj} - R_{ik}).$$

Een analoge formule geldt voor  $\tau(j, i)$  door de rol van  $j$  en  $i$  te verwisselen. Na optellen en gebruik van de symmetrie van de effectieve weerstand, krijgen we

$$\begin{aligned} \tau(i, j) + \tau(j, i) &= (d(i) + d(j))R_{ij} + \\ &\quad + \frac{1}{2} \sum_{k \neq i, j} d(k)(2R_{ij} + R_{kj} - R_{ik} + R_{ki} - R_{jk}) \\ &= \sum_k d(k)R_{ij}. \end{aligned}$$

□

Eerder hadden we nog de slogan over de ‘six degrees of separation’ aangevoerd. Het aantal ‘degrees of separation’ in een graaf is de gemiddelde afstand  $\bar{d}$ . Met experimenten is geprobeerd dit getal te schatten via het versturen van een kettingbrief door een willekeurige groep personen aan een bekend persoon als de Paus.



Hierbij stuurt iedereen die een kettingbrief ontvangt en niet de Paus is, de kettingbrief door naar een willekeurige kennis. Het gemiddelde aantal keren dat de brief wordt doorgestuurd alvorens de Paus te bereiken, wordt als schatting van het ‘aantal degrees of separation’ in het netwerk gebruikt.

**Opdracht 1.4.2.** *Deduceer met behulp van de analyse van de random surfer, of dit een valide experiment is: geeft de uitkomst inderdaad een schatting van het aantal ‘degrees of separation’? U kunt dit formeel doen, of gewoon een klein sociaal netwerkje bouwen en dit doorrekenen!*

## 1.5 Bomen en ‘Matrix-Tree’-stelling

We bespreken nog een laatste interessante interpretatie van de formule

$$R_{ab} = \mathbb{L}_{G,[b]}^{-1}(a, a).$$

Met de regel van Cramer, kunnen we dit herschrijven als

$$R_{a,b} = \frac{\det \mathbb{L}_{G,[a,b]}}{\det \mathbb{L}_{G,[b]}}. \quad (1.6)$$

Hierbij staat “det” voor de determinant van de betreffende matrix.

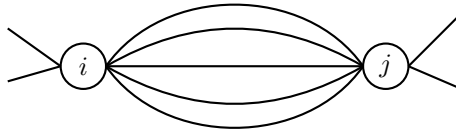
Om de alternatieve interpretatie te begrijpen hebben we het begrip ‘opspannende boom’ nodig. We hebben reeds het begrip ‘boom’ geïntroduceerd: dat is een minimaal samenhangende graaf. Gegeven een samenhangende graaf  $G = (V, E)$ , dan is een opspannende boom een deelgraaf die een boom is.

De volledige graaf met 3 knooppunten de  $K_3$  heeft als opspannende bomen alle deelgrafen met 2 takken. De  $K_4$  heeft in het totaal 16 opspannende bomen.

**Opgave 1.5.1.** *Bepaal alle opspannende bomen van de  $K_4$ .*

De zogenaamde Matrix-Tree stelling specificeert het aantal opspannende bomen van een willekeurige graaf. Er zijn heel veel bewijzen van deze stelling. Het simpelste bewijs, naar mijn weten, is een inductiebewijs.

Hiervoor moeten we de aanwezigheid van parallelle takken toestaan: zij  $m(i, j)$  het aantal takken tussen knooppunt  $i$  en  $j$ . Voor de graad van knooppunt  $i$  krijgen we dan  $d(i) = \sum_{j \in N(i)} m(i, j)$ . Merk op dat  $m(i, j) = m(j, i)$ .



$$m(i, j) = m(j, i) = 5$$

De Laplaciaan van  $G$  wordt dan als volgt:

$$\mathbb{L}_G(i, j) = \begin{cases} d(i), & j = i \\ m(i, j), & j \in N(i). \end{cases}$$

Lussen worden dus niet in aanmerking genomen! We staan wel toe dat de graaf niet samenhangend is (het aantal opspannende bomen is dan 0!).

**Stelling 1.5.1** (Matrix-Tree stelling). *Zij  $G = (V, E)$  een graaf met mogelijk parallelle takken, zonder lussen. Dan is het aantal opspannende bomen van  $G$  gelijk aan  $\det L_{G,[i]}$ ,  $i \in V$ .*

Er is hier dus weer een magische relatie tussen de Laplaciaan die ontstaat door weglaten van een knooppunt. Controle voor de  $K_4$  geeft inderdaad antwoord 16:

$$\det \begin{pmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{pmatrix} = 16.$$

Ik ga ervan uit dat U de regels voor het berekenen van een determinant nog kent! Een direct gevolg van de Matrix-Tree stelling dat we nodig hebben, is het volgende. De correctheid ervan blijkt gemakkelijk uit het bewijs van de deze stelling.

**Gevolg 1.5.2.** *Laat  $(a, b) \in E$ . Dan is het aantal opspannende bomen dat de tak  $(a, b)$  bevat gelijk aan  $\det L_{G,[a,b]}$ .*

**Opgave 1.5.2.** • *Beargumenteer de correctheid van bovenstaande uitspraak.*

- *Laat  $T \subset G$  een deelboom van  $G$  zijn. Bedenk een formule voor het aantal opspannende bomen van  $G$  dat de deelboom  $T$  bevat.*

Alvorens de Matrix-Tree stelling te bewijzen, zullen we bespreken wat het aantal opspannende bomen te maken heeft met de effectieve weerstand. Dat is nu een rechtstreekse invuloefening geworden door vergelijking (1.6), de Matrix-Tree stelling en Gevolg 1.5.2 te combineren:

$$R_{a,b} = \frac{\text{aantal opspannende bomen van } G, \text{ dat de tak } (a,b) \text{ bevat, of, als deze er niet is, na toevoeging van een tak } (a,b)}{\text{totaal aantal opspannende bomen van } G}.$$

Dit is enigszins verrassend! Gaat U deze formule zelf eens na voor de voorbeeldgrafen! Bent U intussen al de door U bedachte formule in Opgave 1.3.1 (4) vergeten? Hier volgt zij expliciet.

**Opgave 1.5.3.** *Bewijs het volgende: als er tussen knooppunten  $a$  en  $b$  maar één keten is, dan geldt dat  $R_{ab} = d_{ab}$ . Met opspannende bomen moet dit gemakkelijk gaan! Geef een intuïtief argument waarom altijd geldt dat  $R_{ab} \leq d_{ab}$ .*

**Bewijs Stelling 1.5.1** We maken gebruik van volledige inductie naar het aantal knooppunten van de graaf, met daarin een geneste inductie naar



het aantal takken. Laat  $n = 2$ . Als de graaf 0 takken heeft, dan is de Laplaciaan een  $2 \times 2$  nulmatrix. Dus  $\det L_{G,[i]} = 0$  voor  $i = 1, 2$ . Het aantal opspannende bomen is ook 0.

Stel er zijn  $m(1, 2)$  takken tussen knooppunten 1 en 2. Dan is

$$L_G = \begin{pmatrix} m(1, 2) & -m(1, 2) \\ -m(2, 1) & m(2, 1) \end{pmatrix}.$$

Het aantal opspannende bomen is gelijk aan  $m(1, 2) = m(2, 1)$ , en dit ook  $\det L_{G,[i]}$ ,  $i = 1, 2$ .

Stel nu dat de bewering waar is voor alle grafen met minder dan  $n$  knooppunten en een willekeurig aantal takken, en voor alle grafen met  $n$  knooppunten en hooguit  $m$  takken. We zullen eerst laten zien dat de bewering ook waar is voor grafen met  $n$  knooppunten en  $m + 1$  takken.

Laat  $G$  zo'n graaf zijn. In het geval dat  $G$  niet samenhangend is, mag U de correctheid van de bewering zelf beargumenteren. Stel dat  $G$  wel samenhangend is.

Kies  $a \in V$  willekeurig. Na eventuele hernummering van de knooppunten mogen we aannemen dat  $a = 1$ . Stel  $(a, b) \in E$  (er is minstens één zo'n tak wegens de samenhang). We mogen ook weer aannemen dat  $b = 2$ . Een opspannende boom van  $G$  kan *geeneen tak*  $(1, 2)$  (geval  $*$ ) of *wel een tak*  $(1, 2)$  (geval  $**$ ) bevatten.

In geval  $*$  kan de inductieveronderstelling meteen worden toegepast. Het aantal opspannende bomen dat geen enkele tak  $(1, 2)$  bevat is gelijk aan  $\det L_{G^*,[1]}$ , waarbij  $G^*$  de graaf is die uit  $G$  ontstaat door de  $m(1, 2)$  takken tussen knooppunten 1 en 2 te verwijderen. De matrix  $L_{G^*}$  heeft de volgende vorm

$$L_{G^*} = \begin{pmatrix} d(1) - m(1, 2) & 0 & \dots \\ 0 & d(2) - m(1, 2) & \dots \\ \vdots & \vdots & \boxed{L_{G,[1,2]}} \end{pmatrix}.$$

In geval  $**$  contraheren we de twee knooppunten 1 en 2 tot 1 knooppunt,  $\bar{v}$  zeg. We construeren een nieuwe graaf  $G^{**} = (V^{**}, E^{**})$  met  $V^{**} = \{v^{**}, 3, \dots, n\}$  en takken

$$(i, j) \in E^{**} \Leftrightarrow \begin{cases} (i, j) \in E, & i, j \in \{3, \dots, n\} \\ (i, 2) \in E, & j = v^{**}, i \in \{3, \dots, n\} \\ (i, 1) \in E & j = v^{**}, i \in \{3, \dots, n\} \end{cases}$$

Gaat U na dat met elke opspannende boom van  $G^{**}$   $m(1, 2)$  opspannende bomen van  $G$  corresponderen. Het aantal opspannende bomen van  $G^{**}$  is

per inductieveronderstelling gelijk aan  $\det L_{G^{**}, [v^{**}]}$ , waarbij

$$L_{G^{**}} = \begin{pmatrix} d(1) + d(2) - 2m(1, 2) & -m(1, 3) - m(2, 3) & \dots \\ -m(1, 3) - m(2, 3) & \boxed{L_{G, [1, 2]}} \\ \vdots & & \end{pmatrix}$$

Nu geldt dat het aantal opspannende bomen van de graaf  $G$  gelijk is aan

$$\begin{aligned} \text{aantal} &= m(1, 2) \cdot \det L_{G^{**}, [v^{**}]} + \det L_{G^*, [1]} \\ &= m(1, 2) \det L_{G, [1, 2]} + \det \begin{pmatrix} d(2) - m(1, 2) & \dots \\ \vdots & \boxed{L_{G, [1, 2]}} \end{pmatrix} \\ &= \det \begin{pmatrix} d(2) & \dots \\ \vdots & \boxed{L_{G, [1, 2]}} \end{pmatrix} = \det L_{G, [1]}. \end{aligned}$$

Dit bewijst de correctheid van de uitspraak voor alle grafen met hooguit  $n$  knooppunten. De inductiestap voor grafen met  $n + 1$  knooppunten volgt analoog: zolang er geen samenhang is, is de uitspraak trivialisierbaar. Met inductie naar het aantal takken volgt het gestelde.  $\square$

**Opmerking** In de laatste twee paragrafen zijn we verder niet ingegaan op generalisaties naar andere ‘weerstand’ op de takken. Deze generalisaties zijn op analoge wijze af te leiden, als U al in enkele opgaven hebt moeten doen. In plaats van werken met de Laplaciaan, moet U de gewogen Laplaciaan gebruiken.

Natuurlijk is een aantal belangrijke vragen onbesproken gebleven: U kunt nu berekenen of het Amsterdamse metronetwerk storingsgevoeliger is dan het Londense. Nou en...?

We kunnen inzicht krijgen in wat een goed gestructureerd netwerk is. Maar uiteindelijk gaat het (wat mij betreft) om het doen van optimalisaties: tussen welke stations levert het aanleggen van een extra lijn een maximale stijging van de robuustheid van het netwerk (gegeven dat ik budget heb voor één lijn). Hierbij zouden allerhande factoren betrokken kunnen worden: de belasting van de verschillende wegtrajecten, de lengtes, etc.

De webgraaf is een gerichte graaf. Voor gerichte grafen is een analoge theorie nog niet bedacht. Omdat niet alle wegen twee-richtingsverkeer toestaan, is het van belang om ook voor gerichte grafen een goede robuustheidsmaat te bedenken. Ook daaraan valt nog het nodige te onderzoeken!

## 1.6 Literatuur

Niet alle bewijzen staan in de literatuur. De auteur vindt het altijd een sport om te proberen bewijzen af te leiden binnen het kader van de behandelde theorie. Dat lukt niet altijd, ook hier zijn nog enkele gaten over gebleven! Die zou U zelf kunnen proberen te dichtten!

Hopelijk hebt U er nog enig plezier in om hier zelf verder ver na te denken, één Uwer leerlingen een profielwerkstuk hierover aan te doen, of in Wiskunde D te behandelen. Er zijn genoeg grappige toepassingen van technieken uit de Lineaire Algebra. Uw vragen zijn welkom!

Wat betreft literatuur: [1] is een masterscriptie met een overzicht van een aantal robuustheidsmaten, en een evaluatie hiervan. Deze is onder mijn begeleiding geschreven. U kunt hierin ook een literatuuroverzicht vinden.

De characterisatie van de effectieve weerstand via de inverse  $L_{G,[b]}$  is betrekkelijk recent, deze staat in [3]. Het elegante bewijs van de Matrix-Tree stelling kunt U vinden in [4].



# Bibliografie

- [1] W. Ellens. Effective Resistance, and other graph measures for network robustness, 2011. Master thesis, University of Leiden, Url <http://math.leidenuiv.nl/nl/theses/239/>. 1.6
- [2] M. Girvan and M.E.J. Newman. Community structure in social and biological networks. *PNAS*, 99(12):7821–7826, 2002. 1.1
- [3] A. Gosh, S. Boyd, and A. Saberi. Minimizing effective resistance of a graph. *SIAM Review*, 50:37–66, 2006. 1.6
- [4] M. Lewin. A generalization of the Matrix-Tree theorem. *Math. Z.*, 181:55–70, 1982. 1.6